

Short Paper 1 - Resubmission

Ethics of T3

Reid McIlroy-Young

October 16, 2016

As of the writing of this report the data set gathered from 1640 of the 2006 incoming students at Harvard College [1] by K. Lewis, *et al.* [2] is not available [3], and it likely will never be available. The moral outrage [4] surrounding the data gathering [2] and successive studies [5] [6] has insured that the data will never again be released. That said, as a researcher interested in social networks if I were given an opportunity to gain access to the *Tastes, Ties, and Time* (T3) I would gladly obtain the dataset. The data are unmatched in length and scope, a pinnacle that will likely never be reached again by public researchers. As I will discuss below, a contemporary institutional review board (IRB) would never allow a study like Lewis's to be run. For this contemporary ethical analysis of T3, I will use the Menlo report's [7] framework which is an update to the Belmont report [8].

Respect for Persons

One of the authors of T3 responded to criticisms [1] of their methods by comparing it to surveillance in a public setting.

Could you require that someone sitting in a public square, observing individuals and taking notes on their behavior, would have to ask those individuals' consent in advance? We have not accessed any information not otherwise available on Facebook. [9]

If the data were truly public, this argument would still not be fully in line with the *Respect for Persons* mandated of the Menlo report. They did not obtain informed consent nor did they attempt to remove those with diminished autonomy.

At the time studies trawling public data on social media was not commonly known to exist so the participants would not have suspected it might be happening, unlike the example given of someone making observations

in public, which is a phenomena in the *Zeitgeist* and thus people can be considered to implicitly giving consent as part of the social contract. By not allowing for even implicit consent the T3 researchers violate the autonomy of the participants. It is also worth noting that they could have obtained consent from every participant without much effort as Harvard College provided them with the email addresses of each of the subjects [2]. They could have contacted the participants and obtained consent, and they did not.

One final note: the authors likely did not know that their ripping techniques were allowing them access to data not intended to be public [1], usage of this data is against the *Respect for Persons* principle as it violates their autonomy, privacy and any semblance of consent.

Beneficence

To the best of my knowledge no harm to any individual can be directly traced to the T3 data gathering or any of the subsequent usage, but, that is due to luck. The authors did not make any real attempts at minimizing risk; all they did to the data is remove names [2], nothing else. In fact they added more information to the set than was available from the scraping of Facebook, they were able to add data about the participants' race, gender and socio-economic background even when it was not given [2]. This data, along with knowledge that the participants are from Harvard College starting in 2006 is enough to de-anonymize a vast majority of them [10]. Thus, if the data were released publicly it would be de-anonymized and its scope and breadth mean it would almost certainly cause some people harm and many embarrassment. The researchers have since ceased distribution of the data but the risk still remains.

We can look at the research done before the lock-down to see if there is some benefit. Wimmer's paper [5] develops our understanding of race and friendship and has over 100

citations [11]. It is the most cited paper building on the T3 data set, the original paper is focused on the new data set so its utility since the data's retraction is miniscule [2] [4]. That said since benefits for the participants are virtually nonexistent and with such a large risk, the ethics of T3 are dubious from a *Beneficence* standpoint since the resulting research has not been worth the risk

Justice

The participants of T3 were not a vulnerable group as a whole, although some individuals may have been (as discussed above). But their efforts in; creating and maintaining their Facebook pages were in no way reimbursed by the study. The researchers used the efforts of 1640 students for their own benefit and a marginal societal benefit, and gave nothing back to the participants.

There are ways they could have provided benefit to the participants even after the publication of the data. The researchers could have given the participants access to monetary reimbursement, provided access to counselling or identity theft protection services to name a few. But they did not, the burden of the study thus was not fairly distributed.

Respect for Law and Public Interest

It appears that the T3 study was done legally and with due diligence. The researchers obtained permission from Facebook, Harvard College and an IRB [2]. They thus thought they were in full compliance with the law; additionally they were very transparent about what they had done even to the point of offering the data set to others. Their actions after publication show at least some effort at accountability, which is admirable, they addressed the criticism [9] and even decided they were wrong, changing their behaviour to minimize harm. Although their intentions were to stay within the bounds of the law, the T3 researches may have violated the law as written and certainly violated the intent when they used RAs in the social network of the target population, allowing them much more than public access.

Conclusion

Since T3 norms around privacy on social media have changed and the chilling effects of the

publication of research and surveillance of social media [12] have been observed. So even if researchers were able to make the same recordings as T3 there would still be significant differences; although, as discussed above the likelihood of this happening is minuscule. Thus the data set is unique and due to its unavailability [3] likely not fully studied. I think that if given the opportunity I would choose to obtain and study the data set but I would not like for it to be made publicly available. My own usage of the T3 data can be defended within the framework of the Menlo Report's principles.

It has been over a decade since the data collection was started so the impact on the individuals by using the data is much lower than the collection 10 years ago. Also there has been time since the T3 article for the participants to learn of and speak out against the report, it is highly likely that at least some are aware of being in the study now. So if I were to put some effort into removing individuals who could be impacted negatively by the data's use I believe I would satisfy the *Respect for Persons* principle. The *Beneficence* principle would similarly be met by this filtering, but there is also a risk of the data being leaked by my usage that should be addressed. I believe I would be able to maintain control of the data set and not cause harm with it, but there is an increase in risk caused by its dissemination that is worth considering. Since the data have already been collected there is no additional work required of the participants so I think the *Justice* principle is satisfied. Similarly the T3 team has already done what was necessary to make the data collection legal so, as long as my own IRB and their obligations were met, I think my obtaining of the T3 data set would be legal and ethical under the *Respect for Law and Public Interest* principle.

References

- [1] Michael Zimmer. On the "anonymity" of the facebook dataset (updated), 2008.
- [2] Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. Tastes, ties, and time: A new social network dataset using facebook.com. *Social Networks*, 30(4):330 – 342, 2008.
- [3] Harvard Dataverse. <https://dataverse.harvard.edu/dataverse/t3>. website, 2016.
- [4] Michael Zimmer. But the data is already public": on the ethics of research in facebook. *Ethics and Information Technology*, 12(4):313–325, 2010.
- [5] Kevin Lewis Andreas Wimmer. Beyond and below racial homophily: Erg models of a friendship network documented on facebook. *American Journal of Sociology*, 116(2):583–642, 2010.
- [6] K. Lewis, M. Gonzalez, and J. Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, 2011.
- [7] D. Dittrich and E. Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Technical report, U.S. Department of Homeland Security, Aug 2012.
- [8] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Belmont report: Ethical principles and guidelines for the protection of human subjects of research, report of the national commission for the protection of human subjects of biomedical and behavioral research. Technical report.
- [9] Jason Kaufman. comment: On the "anonymity" of the facebook dataset (updated), 2008.
- [10] Matthew J. Salganik. *Bit By Bit: Social Research in the Digital Age*. <http://www.bitbybitbook.com/>, 2016. unpublished.
- [11] Web of science.
- [12] Jon Penney. Chilling effects: Online surveillance and wikipedia use. *Berkeley Technology Law Journal*, 31(1):117, 2016.